

# **Exploring the sensitivity of speech-in-noise tests for noise-induced hearing loss.**

For the final published version, please refer to:

Jansen S., Luts H., Dejonckere P., van Wieringen A. & Wouters J. 2013. Exploring the sensitivity of speech-in-noise tests for noise-induced hearing loss. *International journal of audiology*, [epub ahead of print].

Sofie JANSEN<sup>1</sup>, Heleen LUTS<sup>1</sup>, Philippe DEJONCKERE<sup>1,2</sup>, Astrid VAN WIERINGEN<sup>1</sup>, & Jan WOUTERS<sup>1</sup>

<sup>1</sup> KU Leuven – University of Leuven, Department of Neurosciences, ExpORL, Leuven, Belgium

<sup>2</sup> Federal Institute of Occupational Diseases, Brussels, Belgium

## **Key words:**

Speech-in-noise, hearing screening, sensitivity, noise-induced hearing loss

## **Corresponding author:**

Sofie Jansen, ExpORL, KU Leuven

Herestraat 49, bus 721

B-3000 Leuven

BELGIUM

Tel: +32 16 33 04 77

[Sofie.Jansen@med.kuleuven.be](mailto:Sofie.Jansen@med.kuleuven.be)

## Abstract

**Objective:** Recently, the Digit Triplet test was shown to be a sensitive speech-in-noise test for early high-frequency hearing loss in noise-exposed workers. This study investigates if a further improvement is achieved when using a closed set of consonant-vowel-consonant (CVC) speech items with the same vowel, and/or a low-pass (LP) filtered version of the standard speech-shaped noise. **Design:** Speech reception thresholds in noise were gathered for the Digit Triplet, CVC, and CVC\_LP test and compared to the high-frequency pure-tone average (PTA). **Study sample:** 118 noise-exposed workers showing a wide range of high-frequency hearing losses. **Results:** For the 84 Dutch-speaking participants, the CVC test showed an increased measurement error and a decreased between-subject variation, leading to a weaker correlation with the  $PTA_{2,3,4,6}$  ( $R=0.64$ ) and thus a lower sensitivity compared to the Digit Triplet test ( $R=0.86$ ). However, the use of LP-filtered noise resulted in a sensitivity improvement ( $R=0.79$  versus  $R=0.64$ ) due to the large increase in between-subject spread. Similar trends were found for the 34 French-speaking workers. **Conclusions:** Using CVC words with the same vowel could not increase the sensitivity to detect isolated high-frequency hearing loss. With LP-filtered noise, test sensitivity improved, but it did not surpass the original Digit Triplet test.

## Introduction

Extensive exposure to noise – both single exposure to intense sounds as well as long-lasting exposure to more intermediate levels – is known to irreversibly damage the human auditory system. Typically, the outer hair cells (OHC) in the cochlea are affected first, leading to a more broadly tuned and linear basilar membrane response. In terms of functionality, OHC damage leads to (1) an increase in audiometric thresholds, (2) loudness recruitment, and (3) a reduced spectral and temporal resolution (Oxenham & Bacon, 2003; Henry & Heinz, 2012).

In noise-induced hearing loss (NIHL), a shift in audiometric threshold is seen at the high frequencies, with a notch in the 3 to 6 kHz range. With increasing hearing loss, a spread towards the lower frequency regions and 8 kHz will be observed as well (Coles et al, 2000; Niskar et al, 2001; Shargorodsky et al, 2010; Henderson et al, 2011). In this regard, speech intelligibility in quiet, which is related primarily to pure-tone thresholds at 0.5, 1, and 2 kHz, will be unaffected in early NIHL (Glasberg & Moore, 1989; Smoorenburg, 1992; Da Costa, 2001). Complaints about speech understanding in difficult listening situations such as background noise, however, typically arise already at earlier stages. This is also reflected in the rather high correlation ( $R=0.72$ ) between the speech reception threshold (SRT) in noise and the pure-tone average at 2 and 4 kHz ( $PTA_{2,4}$ ), as was found by Smoorenburg (1992) in a group of 200 noise-exposed workers. Similar findings have been reported by Bosman & Smoorenburg (1995). An increase in the supra-threshold SRT in noise and an increase in the pure-tone detection thresholds of frequencies around 4 kHz are thus both early expressions of OHC damage due to noise exposure (Oxenham & Bacon, 2003).

However, Smoorenburg's study also noted that there remained significant unexplained variance in the SRTs of noise-exposed persons based on predictions from the (high-frequency) pure-tone thresholds. This imperfect relation has also extensively been demonstrated in study samples of listeners with a general sensorineural hearing loss (van Rooij & Plomp, 1990; Crandell, 1991; Humes, 2002; Smits et al, 2004; George et al, 2007; Jansen et al, 2012). A first confounding factor is the influence of non-auditory cognitive factors, such as working memory, processing speed, and language skills. These might facilitate top-down processes and with that improve speech comprehension, especially for meaningful everyday sentences (Akeroyd, 2008; Humes, 2007). Furthermore, the SRT-PTA relationship has been shown to be slightly non-linear, with unaffected SRTs in listeners with up to 10 to 15 dB high-frequency hearing loss (Smoorenburg, 1992). A possible reason for this insensitivity might be the high redundancy in meaningful speech items.

Recently, however, Jansen et al (2013b) reported a very strong correlation between the SRT in noise and  $PTA_{2,3,4,6}$  ( $R=0.86$ ) and a high sensitivity and specificity (92 and 89% respectively) to detect very mild high-frequency hearing loss ( $PTA_{2,3,4,6} > 10$  dB HL) in a large group of noise-exposed workers, by using short speech stimuli taken from our most basic vocabulary: digits. By presenting these simple digits between 0 and 9 in random triplet combinations, the cognitive load is considered rather low for the tested person. Furthermore, the lack of context might explain the high sensitivity for the first signs of NIHL. This so-called Digit Triplet test was first developed by Smits et al (2004) as an automated self-test delivered via telephone, and now exists in many languages (Ozimek et al, 2009; Jansen et al, 2010; Vlaming et al, 2011; Meyer et al, 2011; Zokoll et al, 2012; Watson et al, 2012). In all these language versions, a very low measurement error ( $\pm 0.7$  dB) on the adaptively measured SRT was found. Given the general advantages of a supra-threshold speech-in-noise test over a pure-tone threshold measurement for screening purposes, i.e. less affected by ambient noise and less dependent on the absolute presentation level and the transducer frequency response (Plomp & Mimpen, 1979b; Culling et al, 2005), the Digit Triplet test is a highly suitable hearing screening instrument for NIHL.

Since the sensitivity of a speech-in-noise test to detect NIHL proves to be rather dependent on the chosen speech material, the aim of this study was to investigate whether a further improvement in sensitivity could be achieved by adjusting the speech and/or noise material. A first modification was the use of nine monosyllabic words with a consonant-vowel-consonant (CVC) structure, selected from a basic vocabulary, and, most important, all with the same vowel. The rationale for selecting words with the same vowel is that identification of consonants, which typically have their most important information in rapidly changing cues in the higher frequencies, is hypothesized to be more disturbed in listeners with high-frequency hearing loss than the identification of the more steady-state and low-frequency vowels. In this way the test is expected to enlarge the SRT difference between normal-hearing listeners and listeners with mild NIHL.

A second potential adjustment to improve the sensitivity is the use of a low-pass (LP) filtered version of the standard stationary speech-shaped noise, as was first introduced by Leensen et al (2011b). By

attenuating the noise (and thus increasing the SNR) in the high-frequency range, a large gain in SRT is expected for normal-hearing listeners, whereas listeners with NIHL are expected to benefit less or even not at all. Leensen et al (2011b) reported extremely high sensitivity and specificity values of 95 and 98%, respectively, for the detection of mild NIHL. However, due to the participant selection in their study yielding two groups with clearly distinct audiogram patterns, these values are most probably an over-estimation of the real sensitivity and specificity of the test. Therefore, it was still unclear what the benefit would be over the Digit Triplet test in standard speech-shaped noise.

In summary, the objective of this study was to investigate if the sensitivity of a speech-in-noise test for high-frequency hearing loss can further be improved when adjusting the speech material (CVC-words with the same vowel) and/or the noise material (LP-filtered noise) in comparison to the original Digit Triplet test in standard speech-shaped noise. Therefore, these different speech-in-noise tests were presented to a large group of noise-exposed workers. Furthermore, both Dutch- and French-speaking workers were tested, each with speech tests in their own language. In this way, the generalizability to other languages was studied at the same time.

## Methods

### Speech materials

Prior to the evaluation in a noise-exposed population, new speech tests were developed and evaluated in normal-hearing listeners. Details on the development of the French Digit Triplet test are described in Jansen et al (2010) and of the Flemish (i.e. Belgian Dutch) Digit Triplet test in Jansen et al (2013b). For the French and Flemish CVC test, highly similar procedures were used in this study: First, speech items were carefully selected from existing recordings and a stationary speech-shaped noise was generated according to the method described in Versfeld et al (2000). In a second step, optimization measurements were carried out in 10 normal-hearing listeners, by presenting the speech items at a large range of signal-to-noise ratios (SNRs). Based on the intelligibility scores per speech item, the level of each word was adjusted to correct for inter-item SRT differences. This optimization phase is performed in order to get the steepest possible slope of the psychometric function, leading to the lowest possible measurement error for adaptive measurements. In a final step, the adjusted speech material was evaluated in a new group of 10 normal-hearing listeners to establish reference values for the SRT and the slope at the SRT. More details per speech material are given in Table 1 and the reference values for normal-hearing listeners are shown in Table 2.

In contrast to the French Digit Triplet test (Jansen et al, 2010), for which both a telephone version (with limited bandwidth) and a broadband version were developed, the three other tests were optimized and evaluated for broadband use only. As the target of these tests is screening for early signs of high-frequency hearing loss, a higher sensitivity and specificity was expected with broadband stimuli presented through a broadband transducer (Smits et al, 2004; Jansen et al, 2013).

**Table 1.** Details on the selected speech items and presentation format for the Flemish (i.e. Belgian Dutch) and French Digit Triplet and CVC test. (1 Van Wieringen & Wouters (2008); 2 Van Wieringen (personal communication); 3 Fournier (1950); n-AFC = n-alternative forced-choice).

	Digit Triplet test		CVC test	
	Flemish	French	Flemish	French
Speech items	combination of 3 monosyllabic digits:		CVC words with the same vowel:	
	/en/ - 1	/œ/ - 1	/bəl/ - ball	/bak/ - bin
	/twe/ - 2	/dø/ - 2	/ɣət/ - hole	/kan/ - stick
	/dri/ - 3	/trwɑ/ - 3	/həm/ - ham	/fat/ - cat
	/vi:r/ - 4	/katr/ - 4	/kɑr/ - cart	/dal/ - tile
	/veif/ - 5	/sɛk/ - 5	/lɑx/ - laugh	/fam/ - woman
	/zɛs/ - 6	/sis/ - 6	/mɑn/ - man	/mal/ - suitcase
		/set/ - 7	/pɑp/ - porridge	/nap/ - cloth
	/ɑxt/ - 8	/ɥit/ - 8	/tɑk/ - branch	/vaʃ/ - cow
		/noɛf/ - 9	/wɑŋ/ - cheek	/vag/ - wave
Speaker gender	female	female	female	male
Recordings from	LINT <sup>1</sup>	(own recordings)	Lilliput <sup>2</sup>	Listes Mono-syllabiques <sup>3</sup>
Triplet intonation	no	yes	n.a.	n.a.
Announcement words	/	“les numéros”	/	/
Presentation format	10 AFC (digits 0-9)		9 AFC	
Scoring (optimization)	digit		word	
Scoring (final)	triplet		word	

For the CVC test, a LP-filtered version of the standard speech-shaped noise of that test was generated as well. Based on simulations using the speech intelligibility index (ANSI S3.5, 1997), Leensen et al (2011b) found the highest discrimination in SRT between different degrees of high-frequency hearing loss when employing a cut-off frequency of 1.4 kHz and a steep roll-off slope of more than 100 dB per octave. Furthermore, a ‘noise floor’ was added to this LP-filtered noise (original noise attenuated by 15 dB), in order to maintain the advantages of testing in background noise in the high frequencies as much as possible. For our CVC test, the LP-filtered noises were generated with the same characteristics. The reference values for normal-hearing listeners for the CVC test in LP-filtered noise (CVC\_LP) are also shown in Table 2.

**Table 2.** Overview of reference values (SRT and slope at the SRT) for normal-hearing listeners. The Digit Triplet and CVC test were presented at 4 fixed SNRs so that the complete psychometric curve could be fitted, taking into account a chance level of 11.1% for the CVC test. The CVC\_LP test was presented using the simple up-down adaptive procedure whereby no slope estimate is available.

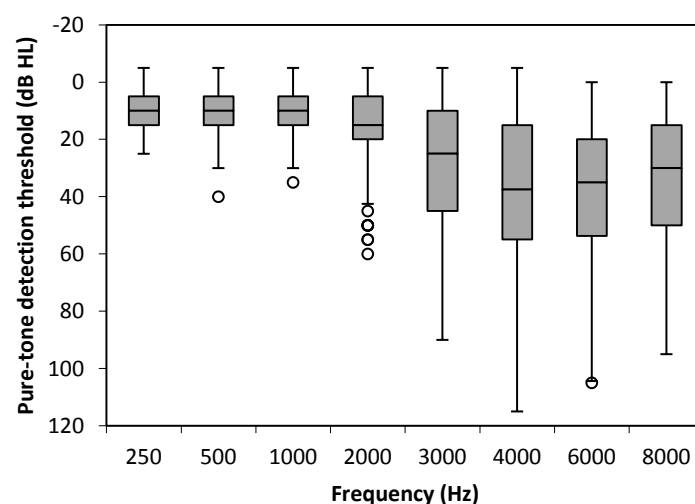
		Digit Triplet	CVC	CVC_LP
SRT (dB SNR) (±SD)	Flemish	-11.7 (±0.6)	-12.0 (±0.6)	-20.6 (±1.7)
	French	-10.5 (±0.3)	-12.1 (±0.7)	-23.8 (±1.9)
Slope (%/dB) (±SD)	Flemish	20.3 (±3.0)	12.7 (±2.6)	/
	French	27.1(±3.0)	13.5 (±1.8)	/

## Participants and test procedure

The same 118 noise-exposed workers that were described in Jansen et al (2013b) participated in this study. There were 84 Dutch- and 34 French-speaking subjects, recruited from five major industrial

companies across Belgium. Two exclusion criteria were a very poor language proficiency both in Dutch and in French, and external or middle ear abnormalities identified through otoscopy and tympanometry. The participants varied largely regarding their age (ranging from 22 to 59), the number of years working in noise, and the type and level of noise to which they were exposed. This resulted in a subject sample with a wide spread of hearing thresholds in the high-frequency range, but with normal or near-normal thresholds up to 1 kHz. Although this high-frequency hearing loss is most probably a mixture of noise-induced and age-related hearing loss in the older participants, the damaging effect of noise exposure was clearly visible in the audiometric notch around 4 kHz. Boxplots of the pure-tone thresholds per frequency for the 118 tested ears are shown in Figure 1.

Each subject completed three SRT measurements in the following test order: (1) The Digit Triplet test in standard speech-shaped noise; (2) the CVC test in standard speech-shaped noise; and (3) the CVC test in LP-filtered noise (CVC\_LP). The Dutch-speaking subjects completed the Flemish version of each test, and the French-speaking subjects the French version. Only one ear was tested and the ear to be tested was chosen randomly, leading to 66 left and 52 right ears. The participants were seated in front of a PC and completed the tests independently, without a test administrator. The signals were sent out via a 24-bit sound card to one side of Sennheiser HDA200 headphones. The level of the noise was constant at 65 dB SPL, and the level of the speech was varied in 2dB steps in a simple up-down adaptive procedure. The noise started 1 s before and ended 500 ms after each trial. For the Digit Triplet test, a list of 27 triplets was presented, and the SRT was defined as the average of the last 22 SNRs. For the CVC and CVC\_LP test, 36 words were presented, and the last 30 SNRs were taken into account for the SRT calculation. The first trial was presented at 0 dB SNR for the two tests in standard speech-shaped noise, and at -10 dB SNR for the test in LP-filtered noise. In this way, most listeners will be able to reach SNRs near their SRT within the first 6 to 7 trials. The software APEX (Francart et al, 2008) was used to play back the speech and the noise at the desired SNR and to automatically score the subjects' responses.



**Figure 1.** Boxplots of the pure-tone thresholds per frequency of the 118 tested ears. The boxes enclose the interquartile range (IQR), the circles represent outliers falling beyond the  $1.5 \times \text{IQR}$  distance from the box, and the whiskers show the minimum and maximum values excluding the outliers.

## Data analysis

In order to evaluate the sensitivity of the three speech-in-noise tests for NIHL, three main parameters were determined and compared in this study. A first factor associated to the general sensitivity of a test is the measurement error – or within-subject test-retest variability. The lower this error, the closer one single measurement outcome will approach a listener's 'real' SRT. The second parameter is the spread of the SRTs across listeners with a wide range of NIHL. In this study, this spread will be characterized by the steepness of the slope of the linear regression line of the SRT versus the high-frequency PTA. As a third parameter, the correlation coefficient of this linear regression analysis, which is related to the steepness of the slope and the measurement error, will be used to quantify the overall sensitivity.

## Results

### Measurement error

The smaller the measurement error of a certain test, the more precisely one can attribute a change (in time) or a difference (between listeners) in SRT to a real change or difference in hearing performance. This measurement error on the adaptively measured SRT can be defined as the quadratic mean of the within-subject standard deviations of repeated measurements. In this study, only one measurement was conducted per test by each person. Therefore, two SRTs were calculated per test, one based on the first half of the trials that were normally taken into account for the SRT calculation, and the other based on the second half. The measurement error was then determined as described above, and divided by  $\sqrt{2}$  to correct for the halving of the number of trials used for the calculation of the two SRTs (Smits et al, 2004; Jansen et al, 2013).

The lowest measurement error was found for the Digit Triplet test, both in Flemish and in French (0.8 dB). The CVC test in standard speech-shaped noise showed a measurement error of 1.1 dB and 1.0 dB for the Flemish and French version, respectively. These values are still in the same order of magnitude as most speech-in-noise tests (e.g. Plomp & Mimpen, 1979a; Nilsson et al, 1994; van Wieringen & Wouters, 2008). However, when presenting the same test in LP-filtered noise, an increase in measurement error was seen, which was more pronounced for the Flemish test (1.6 dB) than for the French test (1.2 dB). A two-tailed paired t-test on the listeners' within-subject standard deviations yielded that the increase in measurement error for the CVC test compared to the Digit Triplet test was significant for the Dutch-speaking listeners ( $p=0.035$ ) but not for the French-speaking participants ( $p=0.093$ ). Also, the increase in within-subject variability from the CVC test to the CVC\_LP test was significant for the Flemish version ( $p=0.001$ ), but again not for the French version ( $p=0.476$ ).

### The SRT-PTA relationship

The main objective of this study was to explore the sensitivity of the three speech-in-noise tests for high-frequency hearing loss in noise-exposed listeners. Therefore, the relationship between the SRT and the  $PTA_{2,3,4,6}$  was investigated. Thresholds at these four frequencies showed the largest effects

of hearing loss which can most probably be attributed to excessive noise exposure. Each of these frequencies separately also proved to be a significant predictor for the Digit Triplet SRT (Jansen et al, 2013). In Figure 2 the scatter plot of the SRT versus the  $PTA_{2,3,4,6}$  is given per speech-in-noise test. To control for the overall difference between the Flemish and French versions of the tests, the SRTs were plotted relative to the reference SRT for normal-hearing listeners (cf. Table 2).

Per test and per language version, a linear regression analysis was performed with the  $PTA_{2,3,4,6}$  as the independent and the SRT as the dependent variable. Two outliers were identified which had an absolute studentized deleted residual larger than  $t[1-\alpha/2n;n-3]$ , with  $\alpha=0.05$ ,  $n=84$  (Dutch-speaking group), and  $n=34$  (French-speaking group). Both outliers occurred for the Digit Triplet test, one for the Flemish and one for the French version. In the left panel of Figure 2, these are the two data points with the highest SRT (15.6 and 15.9 dB SNR relative to the reference). The correlation coefficients and slopes of the linear regression lines, with the outliers excluded, are given in Table 3.

**Table 3. Correlation coefficients and slopes ( $\pm$  standard errors, SE) of the linear regression of the SRT (in dB SNR) versus the  $PTA_{2,3,4,6}$  (in dB HL) with outlying data points excluded.**

	Digit Triplet		CVC		CVC_LP	
	R	Slope ( $\pm$ SE)	R	Slope ( $\pm$ SE)	R	Slope ( $\pm$ SE)
Flemish	0.86	0.12 ( $\pm$ 0.01)	0.64	0.07 ( $\pm$ 0.01)	0.79	0.18 ( $\pm$ 0.02)
French	0.81	0.16 ( $\pm$ 0.02)	0.73	0.12 ( $\pm$ 0.02)	0.91	0.31 ( $\pm$ 0.02)
All	0.85	0.14 ( $\pm$ 0.01)	0.67	0.08 ( $\pm$ 0.01)	0.83	0.22 ( $\pm$ 0.01)

Linear regression analyses were also performed on the complete group of Dutch- and French-speaking subjects together. Only one outlying data point remained for the Digit Triplet test (the French-speaking listener with the highest SRT), whereas a new outlier was found for the CVC test. This was also a French-speaking listener, who had an SRT of 11.6 dB SNR relative to the reference (highest data point in the middle panel of Figure 2). The correlation coefficients and slopes for the linear regression analyses on all subjects together are also given in Table 3.

In general, the CVC\_LP test showed the largest spread of SRTs along the different listeners, with a range of more than 20 dB, resulting in a steep slope of the linear regression line ( $0.18 \pm 0.02$  dBSNR/dBHL for the Flemish and  $0.31 \pm 0.02$  dBSNR/dBHL for the French version). On the other side of the scale, the CVC test in standard speech-shaped noise exhibited a rather shallow slope ( $0.07 \pm 0.01$  dBSNR/dBHL for the Flemish and  $0.12 \pm 0.02$  dBSNR/dBHL for the French version). The slope for the Digit Triplet test had a steepness which was in between the two other tests ( $0.12 \pm 0.01$  dBSNR/dBHL for the Flemish and  $0.16 \pm 0.02$  dBSNR/dBHL for the French version).

For the three Flemish tests, the strongest correlation was seen for the Digit Triplet test ( $R=0.86$ ), which was followed by the CVC\_LP test ( $R=0.79$ ). For the French tests, the CVC\_LP test showed the highest correlation coefficient ( $R=0.91$ ), followed by the Digit Triplet test ( $R=0.81$ ). For both languages, the correlation was the weakest for the CVC test, with  $R=0.64$  for the Flemish and  $R=0.73$  for French version.



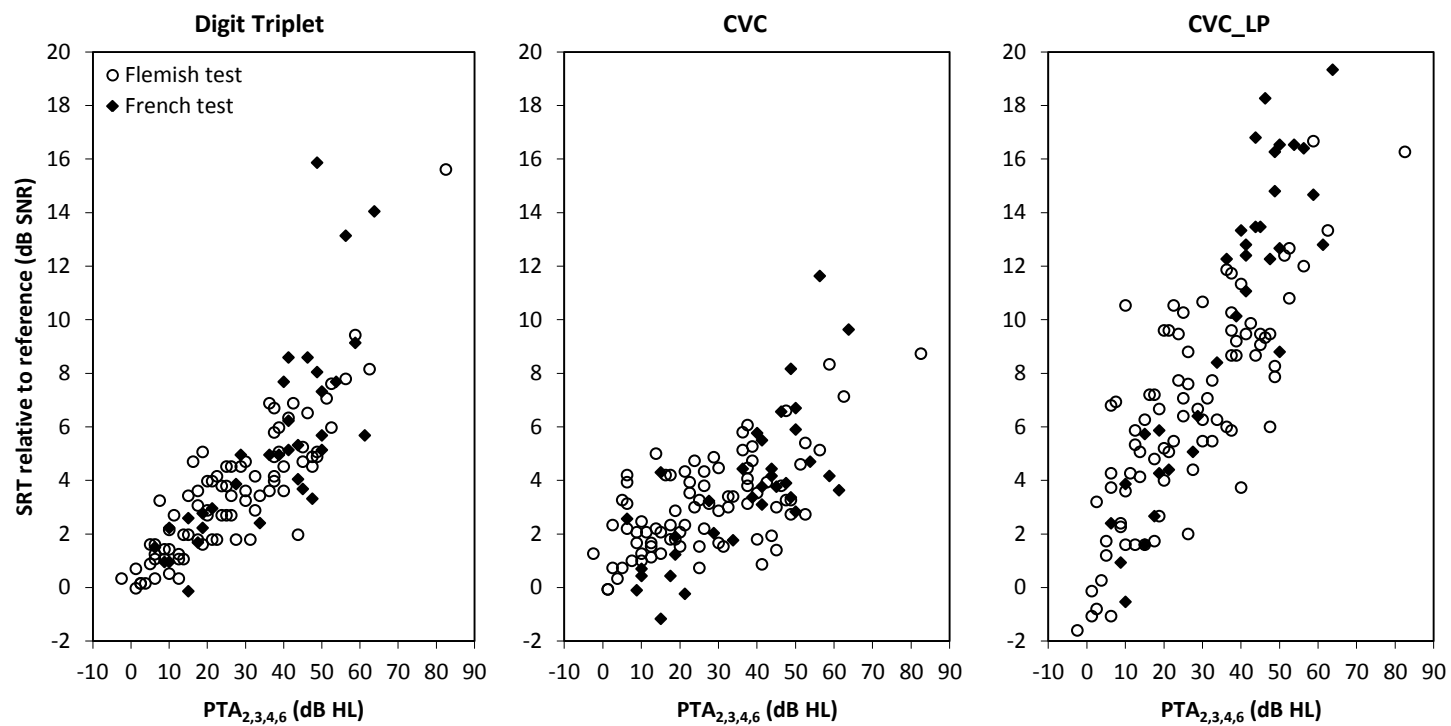


Figure 2. Scatter plot of the SRT (relative to the reference for normal-hearing listeners) versus the  $PTA_{2,3,4,6}$ . Data of the left panel were also shown in Chapter 3.

## Discussion

In general, the results for the two languages show a very similar trend. However, the slopes of the linear regression of the three French tests tend to be steeper than for the Flemish tests. This difference was only statistically significant for the CVC\_LP test, as there was no overlap of the 95% confidence intervals (slope estimate  $\pm 2 \cdot \text{SE}$ ) between the two language versions. Therefore, the results for the two groups will be discussed separately. The results of the largest group of Dutch-speaking participants will be discussed first, each time followed by a (short) discussion of the French results.

### CVC versus Digit Triplet test

For the Flemish tests, the CVC test yielded a significantly shallower slope of the linear regression line compared to the Digit Triplet test (0.07 versus 0.12 dBSNR/dBHL). Together with the significant increase in measurement error (1.1 versus 0.8 dB), the CVC test shows a significantly weaker correlation ( $R=0.64$  versus  $R=0.86$ ; Steiger's  $Z=4.27$ ,  $p<0.001$ ) and thus turns out to have a lower sensitivity for high-frequency hearing loss than the Digit Triplet test. A possible explanation for the higher measurement error of the CVC test is the somewhat shallower slope of the reference psychometric curve of around 13%/dB, whereas the Digit Triplet test has a reference slope of more than 20%/dB (see Table 2 **Error! Reference source not found.**). The steeper the psychometric curve, the more accurately the adaptive procedure will fluctuate around the listener's SRT. Although the number of trials within one adaptive track was higher for the CVC test (36 words versus 27 triplets), this could not compensate for the decrease in precision.

The lack of improvement in SRT spread between the different listeners when using CVC words all with the same vowel is in contrast to what was hypothesized. Possibly, the higher difficulty to identify consonants compared to vowels holds true for normal-hearing listeners equally well as for listeners with high-frequency hearing loss. When comparing the confusion matrices of the Digit Triplet test for the participants with a  $\text{PTA}_{2,3,4,6}$  below 20 dB HL to the participants with a  $\text{PTA}_{2,3,4,6}$  of 20 dB HL or above, this idea seems to be supported. Of all occasions at which the better listeners gave an incorrect answer for the digit /en/, the digit /twe/ was answered 32% of the time, and 36% in the opposite case. Also the digits /dri/ and /vi:r/ were most often mistaken for each other (30% and 31%, respectively). In the hearing-impaired listeners these confusions were somewhat less consistent. Presenting words with the same vowel is thus expected not to give any advantage for normal-hearing listeners over listeners with NIHL.

The same trends were seen for the French tests, though somewhat less strong. The CVC test tended to yield a higher measurement error (1.0 versus 0.8 dB) and a shallower slope of the linear regression line (0.12 versus 0.16 dBSNR/dBHL) and with that a weaker correlation coefficient ( $R=0.73$  versus  $R=0.81$ ) compared to the Digit Triplet test. However, these differences were not statistically significant.

### **CVC\_LP versus CVC test**

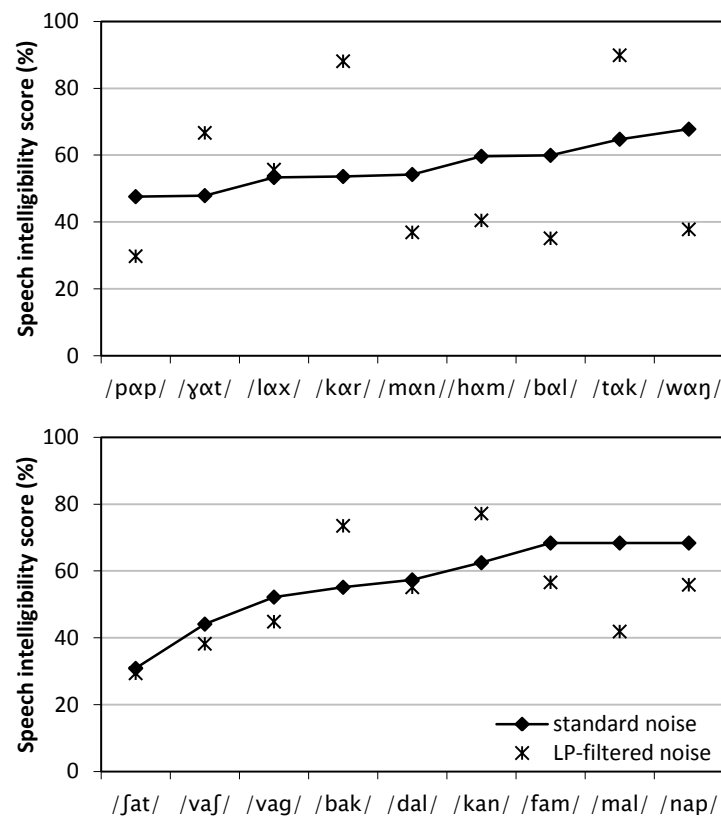
Comparing the CVC\_LP to the CVC test for the Dutch-speaking listeners, a significant increase in the steepness of the slope was seen (0.18 dBSNR/dBHL versus 0.07 dBSNR/dBHL). However, at the same time the measurement error on the SRT proved to be significantly higher as well (1.6 versus 1.1 dB). Still, the CVC\_LP test yielded a significantly higher correlation coefficient ( $R=0.79$  versus  $R=0.64$ , Steiger's  $Z=-2.62$ ,  $p=0.009$ ) and can thus be considered more sensitive than the CVC test in standard speech-shaped noise. For the French-speaking listeners, no significant increase in measurement error was found (1.2 versus 1.0 dB), so that a very clear benefit of the LP-filtered noise was seen ( $R=0.91$  versus  $R=0.73$ , Steiger's  $Z=-3.45$ ,  $p<0.001$ ).

The reason for the increase in measurement error is most probably an increase in inter-item SRT differences when the words are presented in LP-filtered noise. Therefore, the intelligibility scores, averaged over all listeners and all presented SNRs, were determined for each of the 9 CVC words (see Figure 3). In theory, the average score should lie slightly above 50% – the test starts at a relative easy SNR and then converges towards the listeners' SRT – with a small variation across the different words. In Figure 3, this variation is visualized. When comparing the standard deviation across the words when presenting the test in standard speech shaped noise versus LP-filtered noise, a large increase from 7 to 23% was seen for the Flemish test, whereas the French version yielded a much smaller increase from 13 to 16%. Although it is not clear why this effect was larger for the Flemish than for the French version, the increase in measurement error might have been prevented by including a separate optimization phase for the CVC words in LP-filtered noise.

### **CVC\_LP versus Digit Triplet test**

Although there was a clear benefit of using a LP-filtered noise for both language versions of the CVC test, no improvement in sensitivity was seen compared to the original Digit Triplet test in standard speech-shaped noise for the Dutch-speaking participants. The correlation coefficient of the CVC\_LP test was even significantly weaker than for the Digit Triplet test (Steiger's  $Z=1.97$ ,  $p=0.049$ ). This was caused by the loss in sensitivity by using CVC words as speech material. In this regard, the use of digit triplets presented in LP-filtered noise will potentially yield a further increase in sensitivity, and should further be investigated.

To conclude, it needs to be noted that it is questionable whether much higher correlation coefficients between the SRT and the PTA may be achieved. First of all, both psychophysical (behavioral) tests will always yield a certain measurement error. Secondly, it is likely that the supra-threshold identification ability in some listeners will be affected differently (more or less) than their pure-tone detection ability. However, this should be seen as a strength of the SRT, since it measures an ability which is highly relevant for human communication (Shamma, 2011; Ruggles et al, 2011). Persons failing on a speech-in-noise test although passing a pure-tone detection test, might also benefit from an appropriate follow-up so that further hearing loss can be prevented as much as possible and/or an appropriate rehabilitation can be started.



**Figure 3.** Intelligibility scores per CVC word in the standard speech-shaped noise (filled diamonds) and in LP-filtered noise (stars), averaged across all participants and all presented SNRs. Upper panel: Flemish test; Lower panel: French test.

## Conclusions

In this study, the sensitivity of a speech-in-noise test to detect and monitor high-frequency hearing loss in noise-exposed workers, was evaluated with different types of speech and noise materials. Compared to the original Digit Triplet test in stationary speech-shaped noise, it was investigated whether the use of a closed set of CVC words all with the same vowel, and/or the use of a LP-filtered version of the standard noise could improve the sensitivity. Against expectations, the CVC test yielded a lower sensitivity than the Digit Triplet test. This decrease in sensitivity was only significant for the Flemish version of the test, but the French version showed the same tendency. The use of a LP-filtered noise, however, did improve the sensitivity, but it did not surpass the original Digit Triplet test.

## Acknowledgements

The authors would like to thank Lena Vanoirbeek, Annelies Devriendt, and Sophie Radoux for administering the evaluation measurements in normal-hearing listeners for the Digit Triplet and CVC tests. The occupational health physicians Johan De Pooter and Patricia Boulanger (NMBS/SNCB-holding), Ginette Van Walle (Bombardier Transportation), Anne Guyssens (Heijmans Infra NV), Rose-Mary Roussel (FN Herstal), and Jean-Michel Huberlant (ArcelorMittal Liège) are gratefully acknowledged for recruiting the noise-exposed participants of this study. To conclude, we thank the

audiologists Annelies Martens, Christiane Coryn, and Reinharda Van Gool for administering the measurements in the noise-exposed participants.

## References

- Akeroyd M.A. 2008. Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int. J. Audiol.*, 47, S53–71.
- ANSI S3.5 1997. *Methods for Calculation of the Speech Intelligibility Index*. New York: American National Standards Institute.
- Bosman A.J. & Smoorenburg G.F. 1995. Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiol. Off. organ Int. Soc. Audiol.*, 34, 260–84.
- Coles R.R., Lutman M.E. & Buffin J.T. 2000. Guidelines on the diagnosis of noise-induced hearing loss for medicolegal purposes. *Clin. Otolaryngol. Allied Sci.*, 25, 264–73.
- Crandell C.C. 1991. Individual differences in speech recognition ability: implications for hearing aid selection. *Ear Hear.*, 12, 100S–108S.
- Culling J.F., Zhao F. & Stephens D. 2005. The viability of speech-in-noise audiometric screening using domestic audio equipment. *Int. J. Audiol.*, 44, 691–700.
- Da Costa E.A. 2001. Brazilian Portuguese speech material and its application in occupational audiology. *Audiol. Off. organ Int. Soc. Audiol.*, 40, 123–32.
- Fournier J.E. 1950. Vocal audiometry, its technique and results. *Rev. Otoneuroophthalmol.*, 22, 649–56.
- Francart T., van Wieringen A. & Wouters J. 2008. APEX 3: a multi-purpose test platform for auditory psychophysical experiments. *J. Neurosci. Methods*, 172, 283–93.
- George E.L.J., Zekveld A.A., Kramer S.E., Goverts S.T., Festen J.M., et al 2007. Auditory and nonauditory factors affecting speech reception in noise by older listeners. *J. Acoust. Soc. Am.*, 121, 2362–75.
- Glasberg B.R. & Moore B.C.J. 1989. Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scand. Audiol.*, 32, 1–25.
- Henderson E., Testa M.A. & Hartnick C. 2011. Prevalence of noise-induced hearing-threshold shifts and hearing loss among US youths. *Pediatrics*, 127, e39–46.
- Henry K.S. & Heinz M.G. 2012. Diminished temporal coding with sensorineural hearing loss emerges in background noise. *Nat. Neurosci.*, 15, 1362–4.
- Humes L.E. 2002. Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *J. Acoust. Soc. Am.*, 112, 1112–32.

- Humes L.E. 2007. The Contributions of Audibility and Cognitive Factors to the Benefit Provided by Amplified Speech to Older Adults. *J. Am. Acad. Audiol.*, 18, 590–603.
- Jansen S., Luts H., Dejonckere P., van Wieringen A. & Wouters J. 2013. Efficient hearing screening in noise-exposed listeners using the digit triplet test. *Ear Hear.*, 34, 773–8.
- Jansen S., Luts H., Wagener K.C., Frachet B. & Wouters J. 2010. The French digit triplet test: a hearing screening tool for speech intelligibility in noise. *Int. J. Audiol.*, 49, 378–87.
- Jansen S., Luts H., Wagener K.C., Kollmeier B., Del Rio M., et al 2012. Comparison of three types of French speech-in-noise tests: a multi-center study. *Int. J. Audiol.*, 51, 164–73.
- Leensen M.C.J., de Laat J.A.P.M., Snik A.F.M. & Dreschler W.A. 2011. Speech-in-noise screening tests by internet, part 2: improving test sensitivity for noise-induced hearing loss. *Int. J. Audiol.*, 50, 835–48.
- Meyer C., Hickson L., Khan A., Hartley D., Dillon H., et al 2011. Investigation of the actions taken by adults who failed a telephone-based hearing screen. *Ear Hear.*, 32, 720–31.
- Nilsson M., Soli S.D. & Sullivan J.A. 1994. Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.*, 95, 1085–99.
- Niskar A.S., Kieszak S.M., Holmes A.E., Esteban E., Rubin C., et al 2001. Estimated prevalence of noise-induced hearing threshold shifts among children 6 to 19 years of age: the Third National Health and Nutrition Examination Survey, 1988-1994, United States. *Pediatrics*, 108, 40–3.
- Oxenham A.J. & Bacon S.P. 2003. Cochlear compression: perceptual measures and implications for normal and impaired hearing. *Ear Hear.*, 24, 352–66.
- Ozimek E., Kutzner D., Sęk A. & Wicher A. 2009. Development and evaluation of Polish digit triplet test for auditory screening. *Speech Commun.*, 51, 307–16.
- Plomp R. & Mimpen A.M. 1979a. Improving the reliability of testing the speech reception threshold for sentences. *Audiol. Off. organ Int. Soc. Audiol.*, 18, 43–52.
- Plomp R. & Mimpen A.M. 1979b. Speech-reception threshold for sentences as a function of age and noise level. *J. Acoust. Soc. Am.*, 66, 1333–42.
- Ruggles D., Bharadwaj H. & Shinn-Cunningham B.G. 2011. Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proc. Natl. Acad. Sci. U. S. A.*, 108, 15516–21.
- Shamma S.A. 2011. Hearing impairments hidden in normal listeners. *Proc. Natl. Acad. Sci. U. S. A.*, 108, 16139–40.
- Shargorodsky J., Curhan S.G., Curhan G.C. & Eavey R. 2010. Change in prevalence of hearing loss in US adolescents. *JAMA J. Am. Med. Assoc.*, 304, 772–8.
- Smits C., Kapteyn T.S. & Houtgast T. 2004. Development and validation of an automatic speech-in-noise screening test by telephone. *Int. J. Audiol.*, 43, 15–28.

- Smootenburg G.F. 1992. Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *J. Acoust. Soc. Am.*, 91, 421–37.
- Van Rooij J.C. & Plomp R. 1990. Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. *J. Acoust. Soc. Am.*, 88, 2611–24.
- Van Wieringen A. & Wouters J. 2008. LIST and LINT: sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. *Int. J. Audiol.*, 47, 348–55.
- Versfeld N.J., Daalder L., Festen J.M. & Houtgast T. 2000. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J. Acoust. Soc. Am.*, 107, 1671–84.
- Vlaming M.S.M.G., Kollmeier B., Dreschler W.A., Martin R., Wouters J., et al 2011. HearCom: Hearing in the Communication Society. *Acta Acust. united with Acust.*, 97, 175–92.
- Watson C.S., Kidd G.R., Miller J.D., Smits C. & Humes L.E. 2012. Telephone screening tests for functionally impaired hearing: current use in seven countries and development of a US version. *J. Am. Acad. Audiol.*, 23, 757–67.
- Zokoll M.A., Wagener K.C., Brand T., Buschermöhle M. & Kollmeier B. 2012. Internationally comparable screening tests for listening in noise in several European languages: the German digit triplet test as an optimization prototype. *Int. J. Audiol.*, 51, 697–707.